SANDIA REPORT
SAND2014-17042
Unlimited Release
Printed August 2014

Short-Duration Simulations from Measurements

Dean J. Mitchell and Michael Enghauser

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online order:	http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2014-17042
Unlimited Release
Printed August 2014

Short-Duration Simulations from Measurements

Dean J. Mitchell and Michael W. Enghauser, Nuclear Threat Science, 6634
Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185, MS-0782

Abstract

A method is presented that ascribes proper statistical variability to simulations that are derived from longer-duration measurements. This method is applicable to simulations of either real-value or integer-value data. An example is presented that demonstrates the applicability of this technique to the synthesis of gamma-ray spectra.

CONTENTS

FIGURES

1. INTRODUCTION

Simulations provide an efficient method for testing analysis algorithms and evaluating the performance of various types of detectors. Computational methods also enable isolation of variables for parametric studies in ways that may not be feasible when evaluations are performed using only measured data. Although simulation of statistical uncertainties is trivial when the original data does not exhibit statistical variations, uncertainties in the basis data propagate to samples that are derived from measurements. This issue may also pertain to computational methods based on Monte Carlo calculations, which also exhibit statistical uncertainties.

In order to derive representative uncertainty distributions, the true-mean of the basis data must be estimated. Computation of the output from a detector, based on an analytic model, provides one method for estimating the true-mean values without statistical uncertainty. However, this approach has its own shortcomings because the analytic model may exhibit computational errors that exceed statistical uncertainties in the measurement. The basis data may also be smoothed to reduce statistical uncertainties, but smoothing may introduce artifacts that are difficult to quantify. An alternative approach is to use a long-duration measurement as a basis for generating statistical samples with shorter measurement times. Application of this method generally assumes that the long-duration data are representative of true-mean values, and Poisson statistics are applied directly to generate short-duration simulations. However, this approach suffers when statistical samples are generated with measurement times approaching the time basis of the long-duration measurement.

Accordingly, this document describes a method for using measured data to create simulations that reflect the correct statistical uncertainties. Although examples are presented for the synthesis of gamma-ray spectra, the method is suitable for any type of data that exhibits statistical uncertainty consistent with Poisson distributions.

2. THEORY

If true-mean values of the basis data are known, statistical samples can be obtained by extracting random values from a Poisson distribution. The process of obtaining a random sample from this distribution can be represented as the application of a Poisson operator that is based on a true-mean value, μ, as follows:

$$\text{Sample} = \text{Pois}(\mu) \tag{1}$$

The true-mean value is not known precisely if the data is obtained from a measurement. The uncertainty of the true-mean can be reduced by collecting data with a long measurement that is long relative to the simulation. Scaled values of the basis data represent the best estimate for true-mean values associated with shorter measurement times. Accordingly, a statistical sample, y', of the original datum, y, can be approximated as follows:

$$y' = \text{Pois}(sy) \tag{2}$$

where s is a scale parameter that is equal to the ratio of the measurement time for the statistical sample to the measurement time for the original measurement.

Although the impact of statistical uncertainty in the original data is reduced when s is small, uncertainties still contribute to the total variability of the synthetic data. If we assume that an unbiased statistical variability relative to the mean (*p*) can be derived, the statistical sample can be expressed according to Eq. (3).

$$y' = sy + p \tag{3}$$

Let us hypothesize that *p* can be expressed according to the following relationship:

$$p^n = \alpha\,[Pois(sy) - sy]^n \tag{4}$$

Substitution of Eq. (4) into Eq. (3) yields the following:

$$y' = sy + \alpha^{1/n}[Pois(sy) - sy] \tag{5}$$

Equation (5) is identical to Eq. (2) if $\alpha=1$, which is a good approximation when *s* is very small because the statistical uncertainty of the original data is negligible relative to the uncertainty in the statistical sample. Therefore $\alpha=1$ is the proper value in the limit $s<<1$. At the other extreme, no additional uncertainty should be added if the measurement is used to derive a sample with the same measurement time as the original data. Hence, $\alpha=1$ at the limit $s=1$. These limits are achieved if the following relationship pertains:

$$\alpha = 1 - s \tag{6}$$

So Eq. (5) reduces the following representation:

$$y' = sy + (1 - s)^{1/n}[Pois(sy) - sy] \tag{7}$$

The only unknown in Eq. (7) is the value of *n* that produces statistical samples without biasing the data. Insight suggests the proper value of *n* is 2 because statistical uncertainties are normally added in quadrature, but this assumption and the hypothesis that Eq. (4) pertains must be proven. Furthermore, *sy* is only an estimate of the mean as opposed to a true-mean, so derived statistical samples cannot be absolutely correct under all conditions.

3. VALIDATION AND ADJUSTMENTS FOR INTEGER DATA

3.1 Computational Validation

The true-mean of a measurement cannot be known exactly, so a computational approach provides the best way to evaluate the suitability of the derivation in Section 2. In particular, the proper value of parameter *n* in Eq. (7) must be determined. A series of measurements of *y* can be simulated by extracting statistical samples from a Poisson distribution based on the asserted true-mean of the basis data, *t*.

Computation of reduced chi-square (χ_r^2) values provides a means for validating assumptions. The value of χ_r^2, which is defined in Eq. (8), is approximately equal to unity if the data vary according to a Poisson distribution.

$$\chi_r^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{(st - y'_i)^2}{st} \tag{8}$$

The true-mean of shorter-duration simulated data sets is equal to *st*, which appears in the denominator in this expression because it is equal to variance. One million simulations were performed for each of several values of parameters *t* and *s*. Calculations were repeated for $\alpha \equiv 1$ and for values of n equal to 1 and 2. This provided three sets of statistical metrics that can be used to determine the value of *n* and to evaluate the suitability of the proposed sampling approach. Figure 1 compares values of χ_r^2 as a function of *s* and *t* for three assumptions regarding parameters α and *n*.

$\alpha \equiv 1$: Statistical uncertainty in the original data set from which shorter-duration measurements are simulated is neglected with α=1, so values of χ_r^2 tend to exceed unity when the *s* is large.

n=2: The variability of the simulated data is consistent with statistical uncertainties, so the value of χ_r^2 is approximately equal to unity for all values of *t* and s.

n=1: The variability of the simulated data is less than statistical uncertainties.

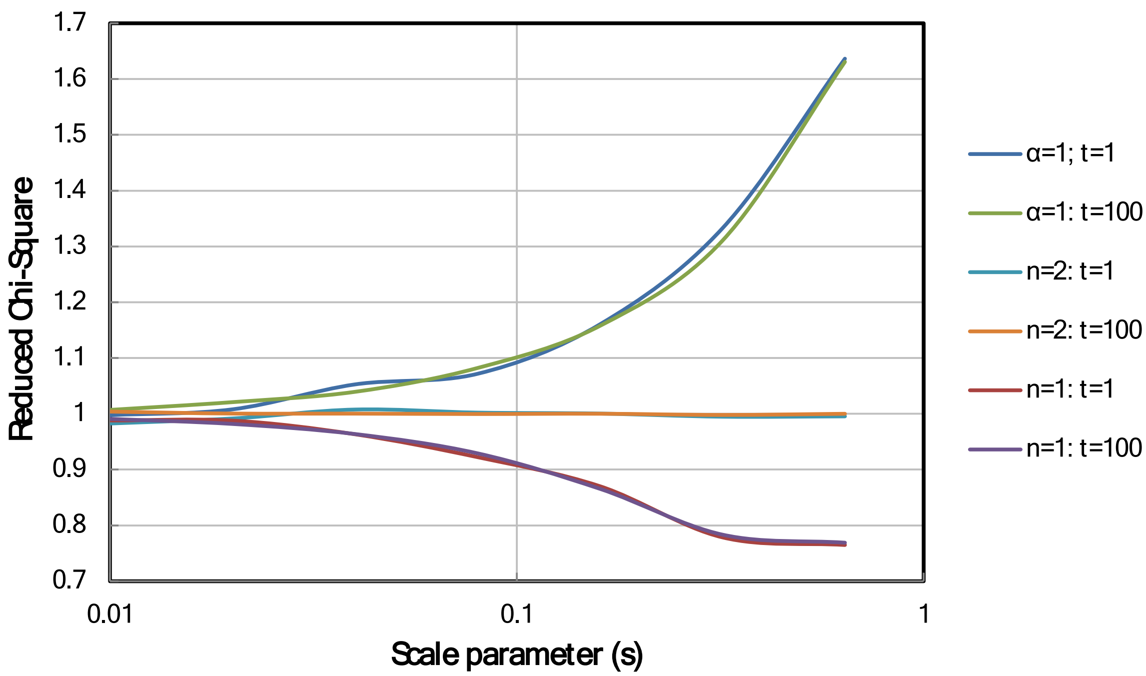

Figure 1. Value of χ_r^2 for the simulated data, *y'*, as a function of the scale parameter, *s*, and the true-mean values of the basis data, *t*.

The statistical metrics presented in Figure 1 show that simulations based on Eq. (7) are consistent with Poisson statistics when n=2. The fact that values of χ_r^2 are approximately equal to unity for all values of *s* and *t* also demonstrates that the underlying assumptions are valid.

3.2 Adjustments for Integer-Value Simulations

The approach described in Section 3.1 is not applicable to all simulations because the values of y' represent floating point numbers whereas measurements often record integer values. The process of rounding data to integer values introduces additional uncertainties that tend to increase variability relative to the mean. This bias can be addressed by reducing the variability of the parameter p in Eq. (3) to compensate for the introduction of round-off errors. The corrected value, p_c, should be related to the statistical component and the component derived from round-off error, σ_r, as follows:

$$p_c^2 = p^2 - \sigma_r^2 \tag{9}$$

This relationship leads to the conclusion that no additional uncertainty is required when $p<\sigma_r$ because the round-off error exceeds the statistical uncertainty. When the average statistical uncertainty exceeds the round-off error (i.e., $sy>\sigma_r$), the ratio of p_c to p can be estimated as follows:

$$\frac{p_c}{p} = \sqrt{\frac{sy - \sigma_r^2}{sy}} \tag{9}$$

When the simulated integer values represented as *Int(y')* are much larger than 1, the average round-off error is equal to 0.25 (this corresponds to the limit $sy>>1$). However, *Int(y')* is generally equal to either 0 or 1 in the limit $sy<<1$, so the round-off errors are either equal to sy or ($1-sy$), respectively. The average probability of obtaining 0 is equal to ($1-sy$) and the average probability of obtaining 1 is equal to sy in this limit. Accordingly, the estimated uncertainty derived from round-off error in the limit $sy<<1$ is given by the following equation:

$$\sigma_r = \sqrt{[sy(1 - sy)]^2 + [(1 - sy)sy]^2} = \sqrt{2}\, sy(1 - sy) \tag{10}$$

Compensation for round-off errors is challenging when $sy \approx 1$ because approximations do not apply for all values of sy. Several methods for estimation of round-off errors in the intermediate regime were explored. The best results were obtained by simply declaring that $\sigma_r = 0.5$ in the range $0.2 < sy < 2$. Although the average round-off error must be less than 0.5, declaring that $\sigma_r = 0.5$ partially compensates for the fact that round-off errors may exceed estimated statistical deviations when $sy \approx 1$, which imposes lower bounds for the deviations that are applied when data are scaled to simulate shorter-duration measurements. Eq. (10) is applied for $sy<0.2$ and $\sigma_r = 0.25$ for $sy>2$.

The dashed curves in Figure 2 show average values of χ_r^2 that are obtained when simulations of y' are rounded to the nearest integer. The solid curves display results that are obtained by the method described in this section for simulating integer data, which produces results that are more consistent with Poisson variability. Even after applying the corrections, values of χ_r^2 tend to exceed unity by a small amount due to deviation minima that are imposed by round-off errors. Round-off errors are negligible relative to statistical uncertainty for large true-mean values, so χ_r^2 is approximately equal to unity regardless of whether the compensation for round-off error is applied.

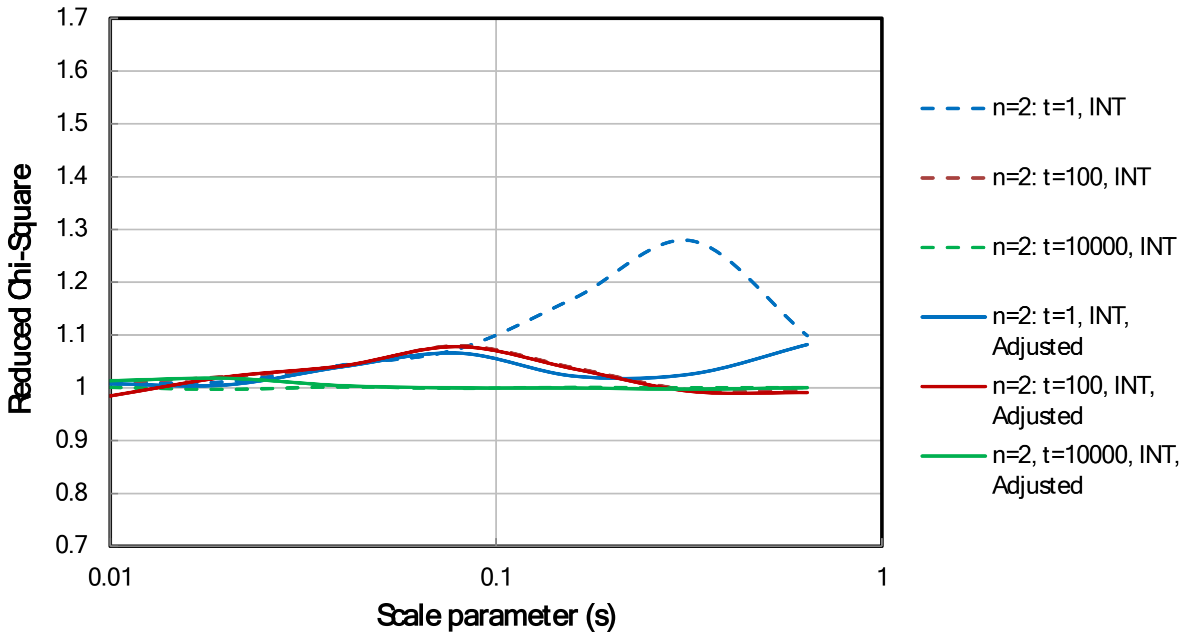

Figure 2. Value of χ_r^2 for simulated integer data, *Int(y')*, as a function of the scale parameter, *s*, and the true-mean values of the basis data, *t*.

4. SYNTHESIS OF GAMMA RAY SPECTRA

The main objective of work described in this document is to enable accurate simulation of gamma-ray spectra based on longer-duration measurements. The simulated spectra should exhibit statistical variability that is consistent with Poisson statistics. As expected, simulations are correlated[1] with the basis data and with each other for values of *s* exceeding about 0.1 because the basis data are applied to represent true-means of the distributions. However, synthesized spectra should exhibit the proper variability with respect to simulations derived from independent basis spectra and spectra that are derived from calculations[1]. Numerous simulations, which were performed as part of this evaluation, demonstrated that χ_r^2 is approximately equal to 1 when spectra are compared. Figure 3 compares two integer spectra that were derived from statistically independent basis spectra. A scale factor of 0.5 was applied to generate these 50-second spectra from two 100-second basis spectra. Values of χ_r^2 ranged from 0.9 to 1.1 in these simulations without an obvious bias relative to a mean value of 1.0.

[1] This is evidenced by values of χ_r^2 that are consistently less than 1.

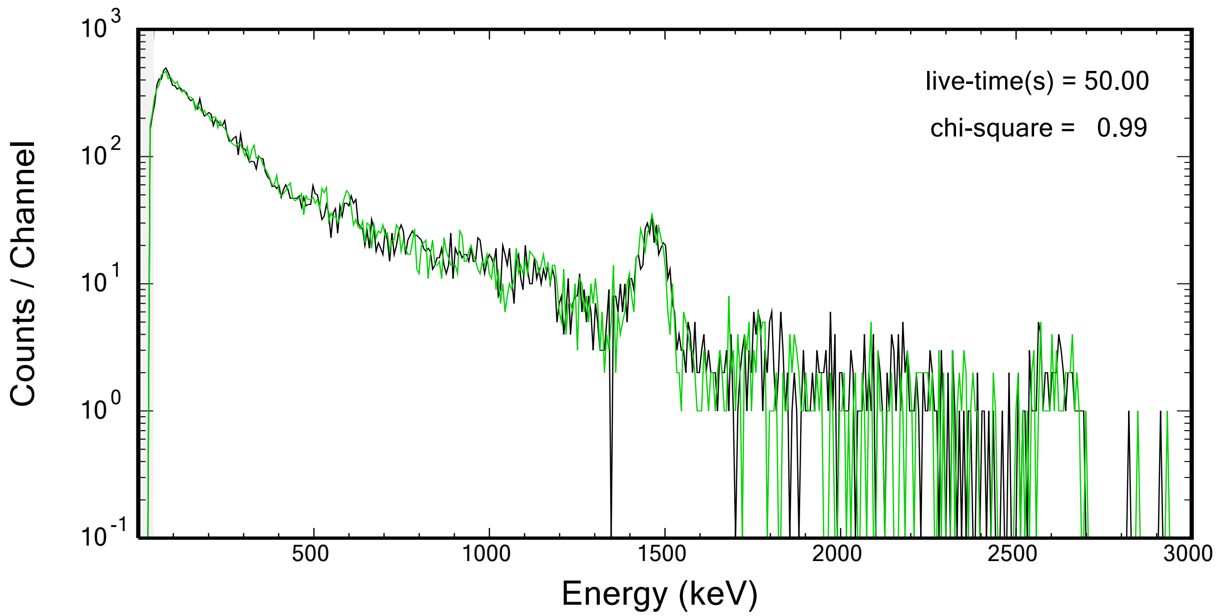

Figure 3. **Simulations that were derived from statistically independent, 100-second background simulations with *s*=0.5 exhibit differences that are consistent with Poisson statistics.**

5. REFERENCES

1. D.J. Mitchell and J. Mattingly, *Rapid Computation of Gamma-ray Spectra for One-Dimensional Source Models (U)*, American Nuclear Society 2008 Annual Meeting (June 2008).

DISTRIBUTION

1	MS0155	C. Hall	6634 (electronic copy)
1	MS0782	D. Mitchell	6634 (electronic copy)
1	MS0782	L. Harding	6634 (electronic copy)
1	MS0899	Technical Library	9536 (electronic copy)

Sandia National Laboratories